

These recent approaches to the prediction of phrasing still do not provide entirely satisfactory results.

According to a first aspect of the present invention, there is provided a method of
5 converting text to speech comprising the steps of:
 receiving an input word sequence in the form of text;
 comparing said input word sequence with each one of a plurality of reference
word sequences provided with phrasing information;
 identifying one or more reference word sequences which most closely match
10 said input word sequence; and
 predicting phrasing for a synthesised spoken version of the input text on the
basis of the phrasing information included with said one or more most closely
matching reference word sequences.

15 By predicting phrasing on the basis of one or more closely matching reference word
sequences, sentences are given a more natural-sounding phrasing than has hitherto
been the case.

Preferably, the method involves the matching of syntactic characteristics of words or
20 groups of words. It could instead involve the matching of the words themselves, but
that would require a large amount of storage and processing power. Alternatively, the
method could compare the role of the words in the sentence – i.e. it could identify
words or groups of words as the subject, verb or object of a sentence etc. and then
look for one or more reference sentences with a similar pattern of subject, verb,
25 object etc.

Preferably, the method further comprises the step of identifying clusters of words in
the input text which are unlikely to include prosodic phrase boundaries. In this case,
the reference sentences are further provided with information identifying such
30 clusters of words within them. The comparison step then comprises a plurality of
per-cluster comparisons.

By limiting the possible locations of phrase boundary sites to locations between clusters of words, the amount of processing required is lower than would be required were every inter-word location to be considered. Nevertheless, other embodiments are possible in which a per-word comparison is used.

5

Measures of similarity between the input clusters and reference clusters which might be used include:

a) measures of similarity in the syntactic characteristics of the input cluster and the
10 reference cluster;

b) measures of similarity in the syntactic characteristics of the words in the input cluster and the words in the reference cluster; and

15 c) measures of similarity in the number of words or syllables in the input cluster and the reference cluster.

d) measures of similarity in the role (e.g. subject, verb, object) of the input cluster and the reference cluster;

20

e) measures of similarity in the role of the words in the input cluster and the reference cluster;

f) measures of similarity in word grouping information, such as the start and end of
25 sentences and paragraphs; and

g) measures of similarity in whether new or previously information is being presented in the cluster.

30 One or a weighted combination of the above measures might be used. Other possible inter-cluster similarity measures will occur to those skilled in the art.

In some embodiments, the comparison comprises measuring the similarity in the positions of prosodic boundaries previously predicted for the input sentence and the positions of the prosodic boundaries in the reference sequences. In a preferred embodiment a weighted combination of all the above measures is used.

5

According to a second aspect of the present invention, there is provided a text to speech conversion apparatus comprising:

a word sequence store storing a plurality of reference word sequences which are provided with prosodic boundary information;

10 a program store storing a program;

a processor in communication with said program store and the word sequence store;

means for receiving an input word sequence in the form of text;

wherein said program controls said processor to:

15 compare said input word sequence with each one of a plurality of said reference word sequences;

identify one or more reference word sequences which most closely match said input word sequence; and

20 derive prosodic boundary information for the input text on the basis of the prosodic boundary information included with said one or more most closely matching reference word sequences.

According to a third aspect of the present invention, there is provided a program storage device readable by a computer, said device embodying computer readable
25 code executable by the computer to perform a method according to the first aspect of the present invention.

According to a fourth aspect of the present invention, there is provided a signal embodying computer executable code for loading into a computer for the
30 performance of the method according to the first aspect of the present invention.